Indirect assessment of practical science skills: Development and application of a taxonomy of written questions about practical work

Frances Wilson OCR, **Stuart Shaw** Cambridge Assessment International Education,
Neil Wade OCR, **Sarah Hughes** OCR, and **Sarah Mattey** Research Division
(The study was completed when the fourth and fifth authors were based in Cambridge Assessment International Education)

Background

Practical work is central to science education and is used not only to support the development of conceptual knowledge, but to enable students to develop a wide range of skills, including data handling, experimental design and equipment manipulation (Wilson, Wade, & Evans, 2016). Practical work may be assessed using many different forms of assessment, including coursework projects, practical exams and written questions in exams. Different forms of assessment may assess different aspects of this complex domain. As such, it is important to establish a clear understanding of the skills and knowledge which are assessed by each form of assessment. In this article, we focus on the evaluation of one method for the assessment of practical science: written exam questions.

Abrahams and Reiss (2012) define practical work as "an overarching term that refers to *any* type of science teaching and learning activity in which students, either working individually or in small groups, are involved in manipulating and/or observing real objects and materials" (p.1036). This is as opposed to virtual materials, such as those obtained from a video or digital simulation, or even from a text-based account. Furthermore, Lunetta, Hofstein, and Clough (2007, p.394), suggest that practical activities may:

1. be experienced in school settings where students interact with materials to observe and understand the natural world;
2. be individual or group, or large group demonstrations;
3. vary on a continuum between highly structured and teacher-centred through to open inquiry;
4. last several weeks, including outside school activities, or less than 20 minutes;
5. use a high level of instrumentation or none at all; and
6. include activities where students gather data to illustrate a principle/relationship (deductive), and those where students gather data and try to work out patterns or relationships from the data (inductive).

Practical work may therefore span a wide range of different activities, which may be used by teachers to serve many different learning aims. Given the impact of assessment on what is taught and learned, it is important to understand exactly what skills are assessed, and how they relate to the specified aims of practical work within the curriculum. Previous work in this area has articulated multiple purposes for practical science, in varying levels of granularity, and based on different interpretations of what is considered to be practical work. For example, focusing primarily on higher education (HE), Kirschner and Meester (1988) catalogued 120 different aims for practical work, which they synthesised into 8 overarching aims. However, these aims are not given equal importance by teachers. Additionally, the relative importance of the aims of practical work at different stages of education and across the different science subjects may vary (Kerr, 1963).

Internationally, a wide range of forms of assessment are used to assess practical skills. Abrahams, Reiss, and Sharpe (2013) distinguish between the *Direct Assessment of Practical Work*, which includes a practical exam, where students are observed carrying out a practical activity, and *Indirect Assessment of Practical Work*, where students may be assessed on the product of a practical activity (e.g., a laboratory report), or may be asked written questions in an exam. Although written questions about practical work cannot assess students' ability to manipulate equipment, they are frequently used to assess other aspects of practical work, including knowledge of experimental procedures and techniques, data analysis and presentation, and the interpretation of data with respect to scientific concepts. A secondary aim of the inclusion of written questions about practical work is to encourage the teaching and learning of practical work, because this will serve as a good preparation for the assessment. This will support the development of those skills (e.g., manipulative skills) which are not directly assessed. Written questions may be used as part of a written exam, or comprise a whole exam paper. Given the diversity of practical science skills which might be assessed, it is important to understand which skills are assessed and in what proportion in any given assessment. Teachers and students need to know how to structure teaching and learning to ensure that they are prepared for the assessment. Similarly, assessment organisations (AOs) that set the assessments must ensure that the balance of skills assessed is appropriate for the intended curriculum, both within a single assessment, and across the lifetime of a qualification.

Although individual syllabuses and curricula state the skills and knowledge which should be assessed in a specific qualification (e.g., the assessment of how observations are recorded, measured and estimated), currently there is no framework for categorising the skills and knowledge assessed using written questions about practical work which has been designed to be used to compare different qualifications, and used in different contexts.

This article describes a study in which the development of a taxonomy is first described, then its application in evaluating current science qualifications explored. The taxonomy aims to classify practical skills assessed using written questions about practical work. The taxonomy seeks to provide an accessible description of the skills, knowledge, and understanding (constructs) which underlie practical science in written exam papers. A taxonomy of practical science skills on questions about practical science has the potential to allow evaluation and monitoring of practical science questions for AOs as it allows for comparisons of skills assessed over time, between papers and between subjects. Cambridge Assessment International Education assessments include written practical science questions (as an alternative to practical exams) and so an evaluation as to how the skills assessed have changed over time, and how they vary between subjects, would provide additional information on the performance of the assessments. OCR assessments of written practical science were introduced in 2016 as a result of reforms in the qualifications. Therefore, evaluating whether the assessments are similar to Sample Assessment Materials (SAMs), and how subjects compare to one another in terms of the skills assessed, would aid the meeting of regulatory requirements to evaluate and monitor the new assessments.

The study described here was undertaken in two phases:

Phase 1: Development of the taxonomy aimed first to develop a taxonomy, then to elicit feedback from stakeholders in order to determine its efficacy. It, therefore, entailed two stages:

1. Teacher and stakeholder input; and
2. Refining the taxonomy.

Phase 2: Application of the taxonomy had three aims:

1. To analyse the skills assessed in different papers, for different sciences and across several years;
2. To use data on the functioning of individual questions to evaluate the quality of questions assessing particular skills; and
3. To determine in what other contexts and for what benefits the taxonomy could be applied.

Phase 1: Development of the taxonomy

The focus of this study was General Certificate of Secondary Education (GCSE), General Certificate of Education (GCE) Advanced Level (A Level) and Advanced Subsidiary Level (AS Level) science question papers from one UK AO, and an international GCSE (IGCSE) written alternative to practical science question papers.

For the first phase of research, the papers analysed included:

- IGCSE *Alternative to Practical* (ATP) written question papers from June 2014 and 2015 for Biology, Chemistry and Physics. The ATP papers include experimental contexts covered by the practical assessment of the IGCSE science assessments, but are indirect assessment of practical skills; and
- GCSE and GCE SAMs for each of the three science subjects (designed to reflect recent reforms in the assessment of practical science). The questions in these papers are referred to as *embedded IAPS* (Indirect Assessment of Practical Skills).

Stage 1: Teacher and stakeholder input

Teachers were asked to categorise practical questions from science papers. Teacher-centred input was considered to be crucial, given concerns about the impact of assessing science practical work through written questions on classroom practical work. A group of nine science teachers was recruited (three Biology, three Chemistry and three Physics). Teachers had differing experience of teaching GCE AS Level, GCSE, and IGCSE.

Stage 1 comprised two sessions: a *subject group* session, and a *cross-subject group* session.

Subject-group session

Teachers were required to work in groups of threes, in subject specialisms (e.g., three Biology specialists working together). Each group was provided with packs of question papers appropriate to their specialism, with questions about practical work highlighted. Teachers were asked to think about how they would categorise the highlighted questions, in terms of the type of knowledge or skill that was being assessed, and to comment on any aspects relating to progression from GCSE and IGCSE to AS Level. Having decided on a list of categories, each group was then required to generate a short description of the category, such that another science teacher could use their categories to identify questions which fell into each category. Teachers were also asked to list any questions in their packs of papers that fell into each category. Finally, teachers were encouraged to log any issues or challenges arising whilst they worked, especially in relation to questions that were difficult to categorise.

Cross-subject group session

In a subsequent session, teachers were asked to work in groups of three, with one subject specialist from each subject comprising each group. This time, groups were asked to review the three category lists developed in the first session, and identify areas of convergence and divergence. Additionally, each group was encouraged to consider the questions identified for each category in the first session. The cross-subject session culminated in a plenary attempt to develop one set of generalised categories which could be used for all three science subjects. The set of categories helped inform the construction of a draft taxonomy of skills from written questions about practical work.

The workshop held with teachers (Stage 1) produced lists containing a total of 57 words and phrases; the Biology subject group produced a list of 23 words and phrases, Chemistry 14 and Physics 20 with very limited direct repetition. Separating individual actions from phrases which combined several skills allowed the creation of a combined set of 48 categories, which was subsequently distilled into 15 statements.

Stage 2: Refining the taxonomy

Stage 2 consisted of eliciting feedback on the distilled taxonomy (constructed in Stage 1) from stakeholder groups (including representatives from HE, subject associations, and teachers). These groups identified the need for segregation of drawing skills into two separate categories. Firstly, the accurate representation (and labelling) of objects observed, such as required in Biology or Geology and, secondly, the more abstract diagrammatic representation of objects using defined symbols, such as electrical circuits, molecular structures, or laboratory apparatus.

The distilled taxonomy developed in Stage 1 was also scrutinised by assessment specialists in each subject to categorise assessment items for a subsequent qualitative analysis. In carrying out this activity, any difficulties identified in assigning a category to a question or item were catalogued.

Table 1: Description of the taxonomy of practical science skills

Skill	*Brief explanation*	*Detailed explanation*
Diagrams (apparatus and circuits)	Representation of equipment or circuits using accepted symbols.	The ability to represent circuits or apparatus in accepted forms.
Drawing (Biology) and labelling	Biological drawing, accurate representation.	The ability to accurately represent objects observed and to appropriately label them.
Making measurements	Reading a scale from a diagram (prevalent in alternative to practical papers).	A practical test of the ability to use a scale to take a reading. In more complex scenarios, a time-lapse photograph or scale diagram may be reproduced requiring students to make measurements.
Recall	Application of knowledge.	A response to a question relating to a practical activity defined in the specification, drawing on recall of theory, or carrying out the activity.
Capturing data	Observing and reading data, interpreting data from a table or graph, recording data.	This is distinct from making measurements. In this case, the data is taken from a table or graph. As opposed to plotting a graph, in this instance the student retrieves data from the graph. The skill is also deemed to involve the ability to record the data appropriately.
Data analysis	Problem solving.	Having identified the appropriate data from a range of sources or different calculations, linking the information to allow the solution of a problem.
Data handling	Calculation (e.g., calculating gradient).	The use of data to carry out a calculation using a formula which may require rearrangement, or the linking of formulae.
Data interpretation/ Identifying trends	Plotting and interpreting graphs.	Transferring data from tabular to graphical format. Being able to identify trends from the graph (telling the story of the graph).
Data quality	Evaluation of data and conclusions.	Comparing the outcome of an activity with the anticipated or accepted outcome. Using numeric processes to comment on the quality of the data, with possible reference to the uncertainty of the process or measurements.
Experimental design	Method, planning and procedure including identification of variables.	An understanding of the processes involved allowing the identification of variables and the ability to propose an experiment to demonstrate the required hypothesis or outcome.
Predicting outcomes	Understanding processes.	Using knowledge and understanding of a process to anticipate the likely outcome of a given sequence of events. Often examined by giving a scenario for a practical activity, specifying a change in the circumstances, and asking for identification of any changes in the outcome.
Use of apparatus and techniques	Application of knowledge of practical skills.	Typical questions could be the evaluation of the use of specific equipment or suggestion as to possible improvements. This has implied understanding rather than straight recall.

Overall feedback from stakeholders, as well as comments on the taxonomy collected during a *Science Forum* organised by a UK AO, contributed to the refinement of the statements and explanations to produce the final taxonomy. Assessment specialists/forum participants contributed to a number of adjustments to the taxonomy. This was particularly important as the taxonomy was different in nature to earlier categorisations of practical work depicted in the literature: It differed against such headings as planning, observing, analysing and evaluating – labels which describe the nature of a question, rather than the skills which may be used in its completion. This review of the taxonomy, with a range of stakeholders, teachers, university lecturers, and representatives of the scientific community (via the Science Forum), also gave feedback leading to the addition of the abstract diagrammatic representation of electrical circuits and stylised two dimensional diagrams of chemical apparatus as a second separate drawing skill.

The final taxonomy is shown as Table 1.

Three key issues were raised by the activities to this point:

1. The segregation of activities related to data into four distinct categories;
2. The distinction between drawing an accurate representation and the use of symbols; and
3. The nature of the taxonomy.

Teachers had clearly identified a range of different skills pertinent to the use of data. Consequently, we were able to identify questions which linked directly to these skills. Examples of the categories can be demonstrated in the following questions:

Example 1: Capturing data/Data handling

This 4-mark item (shown in Figure 1) requires (i) that the candidate captures the appropriate data from the graph, and (ii) that they handle the data to calculate the gradient and hence determine the Young modulus of the metal.

23 The extension of a metal wire is x when the tension in the wire is F. The table in **Fig. 23.1** shows the results from an experiment, including the stress and the strain values.

F / N	x / 10^{-3} m	stress / 10^{7} Pa	strain / 10^{-3}
1.9	0.4	1.73	0.20
4.0	0.8	3.50	0.40
5.9	1.2	5.21	0.60
8.0		7.00	0.80
9.0	1.8	7.95	0.90

Fig. 23.1

(a) Complete the table by determining the extension when the tension is 8.0 N. **[1]**

(b) **Fig. 23.2** shows a graph of stress against strain for the metal.

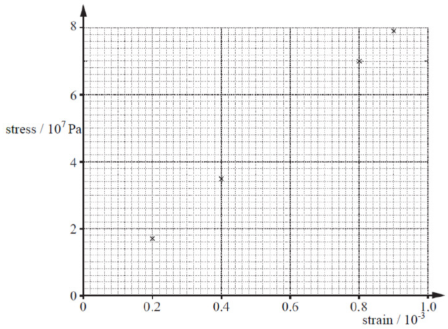


Fig. 23.2

(i) On **Fig. 23.2**, plot the data point corresponding to the tension of 5.9 N and draw the line of best fit through all the data points. **[1]**

(ii) Use **Fig. 23.2** to determine the Young modulus of the metal.

Young modulus = ……………………Pa **[2]**

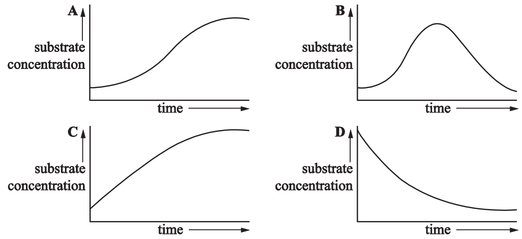
Figure 1: Capturing data/Data handling

Example 2: Data interpretation/Identifying trends

The question in the next example (shown as Figure 2) is typical of those looking to assess the understanding of a practical activity by requiring the candidate to identify the trend of a graph to match the defined scenario. The other main question type in the data interpretation category is in transferring tabular data to graphical form, as incorporated in section (b)(i) of Example 1 (shown in Figure 1).

10 A group of students monitored the **substrate** concentration during an enzyme-controlled reaction.

Select the graph that correctly shows how the substrate concentration changes during the course of the reaction.



Your answer **[1]**

Figure 2: Data interpretation/Identifying trends

Example 3: Data quality

This is an open-ended question (shown in Figure 3) with a total of 6 marks available. It incorporates a range of lower order items from the taxonomy, which then contribute to the higher order. The question provides significant data and uses earlier questions relating to the data and graphs provided to subsequently contribute to the higher demand in making a comparison between the two possible methods.

(iii)* The student calculated the total amount of charge to flow from the capacitor in the first 30 seconds. She used two methods:

Method 1 – using the relationship $\Delta Q = C\,\Delta V$

Method 2 – using $I = \frac{V}{R}$ *and* the area under the I (t) graph where R is 5.0 kΩ

Show how the total charge flow may be calculated in both cases and comment on how the two methods compare. **[6]**

Figure 3: Data quality

Example 4: Drawing (Biology)

The review of an early iteration of the taxonomy identified the skills relating to drawing, evaluation and explanation as missing items. Quillin and Thomas (2015) discuss how drawings vary in the extent to which they are representational or abstract, both within Biology and the wider sciences. The following question (shown in Figure 4) depicts a biological drawing as an accurate representation of an object.

4 As the human population continues to grow there is an ever increasing need to increase food production.

(a) Alfalfa is grown mainly for animal feed as it is rich in protein, minerals and vitamins. The leaves can also be used as a dietary supplement in human nutrition.

Fig. 4.1 below shows the transverse section of an alfalfa leaf.

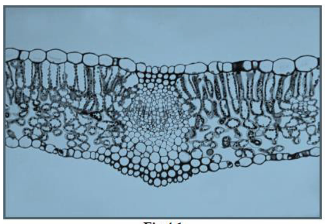


Fig 4.1

In the space below draw a labelled and annotated low power plan of the transverse section of the alfalfa leaf shown in **Fig. 4.1**.

[4]

Figure 4: Drawing (Biology)

Phase 2: Application of the taxonomy

For the second phase of research, a separate set of papers (those taken by students in 2016) was used. This set of papers was the first assessment of the redeveloped international syllabuses, though there were no changes to the assessment of practical skills. Thus, the addition of the 2016 question papers allowed analysis of current UK and international assessment models.

The application phase sought to investigate whether the taxonomy could be applied to items on written practical science within past question papers and SAMs. The analysis, using the taxonomy, aimed to demonstrate how the taxonomy could be used to evaluate and monitor science practical questions by addressing two research questions:

1. *Are the skills in the taxonomy assessed?*

   This question was addressed by comparing the coverage of skills assessed in IGCSE, GCSE, and SAM items on question papers across years and subjects. This use of the taxonomy allows for AOs to monitor and evaluate whether assessments are consistent in assessing practical science skills.

2. *How do questions addressing particular skills perform?*

   This question was addressed by comparing whether particular skills from the taxonomy are associated with particular characteristics of item performance. This use of the taxonomy allows for AOs to evaluate whether particular skills have different characteristics in terms of difficulty and suitability of the item within the rest of the question paper. For instance, if certain skills have higher difficulty values than others, AOs may use this information to train question setters to recognise how difficult particular skills are, or to determine how skills should be assessed in the future. These analyses also provide evidence that the skills being assessed are appropriate to the assessment, in that they do not lead to high omission rates, or have a higher than anticipated difficulty.

Research question 1: Are the skills in the taxonomy assessed?

In order to address this question, subject specialists from both OCR and Cambridge Assessment International Education reviewed the question papers listed below. They first reviewed each item and judged whether it assessed a practical science skill, and if so, they then identified which skills were assessed using the taxonomy. Multiple skills could be assigned for each item, and the first skill assigned was used as the primary skill that the item assessed.

- IGCSE *Alternative to Practical* (ATP) written question papers from June 2014 – 2016 for Biology and Chemistry, and June 2013, 2015 and 2016 for Physics. These are referred to as ATP questions.
- GCE AS Level written question papers for June 2016 for each of the three science subjects (both specifications A and B for Biology and Chemistry, and Specification B for Physics). In addition, SAMs for Biology and Chemistry were also used. These are referred to as *embedded* IAPS questions.

As the subject specialists in this phase applied the taxonomy to items, they found them to be assessing practical science skills in past question

Figure 5: Practical skills assessed across the three sciences

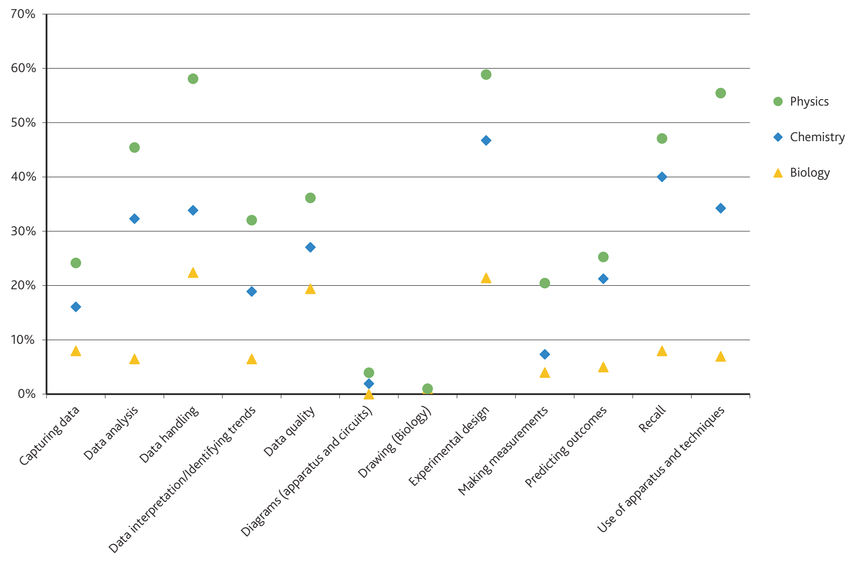
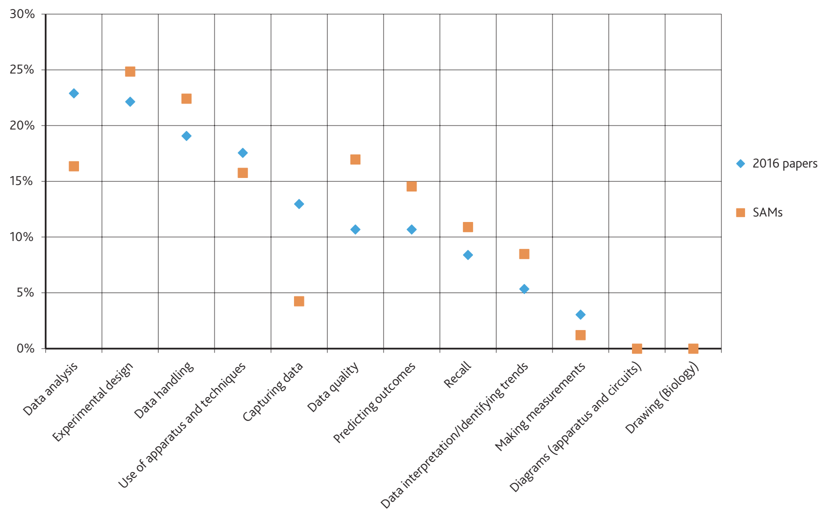


Figure 6: Practical skills assessed across the three sciences (SAMs and live papers)



papers, and analyses were carried out on the items identified. Firstly, the number of each practical skill assessed from all question papers was counted and comparisons made across years, AOs, and subjects. In addition to the planned comparisons, additional comparisons were explored in terms of comparing the mark tariffs of items assessing different skills, as a form of additional information that could be used to evaluate the performance of the assessments. Secondly, item level data was collected for all question papers investigated, and used in conjunction with the skills assigned to each item to compare performance measures of items according to the different skills they assessed.

All of the skills in the taxonomy were judged to be assessed to some extent on each paper analysed, although to varying degrees (see Figure 5). The most commonly assessed skills were *Experimental design, Data handling, Use of apparatus and techniques* and *Recall*. The popularity of Recall is probably impacted by the fact that half of the questions that assess this skill are also assessing other skills. The assessment of the very specific skills of *Diagrams (apparatus and circuits)* and *Drawing (Biology)* were rare.

In Figure 5, we show a comparison of the skills assessed by the three science subjects Physics, Chemistry, and Biology.

The proportion in which skills are assessed in the SAMs and the live 2016 papers are compared in Figure 6. This allows comparison of the 2016 live assessment against design criteria exemplified in the SAMs. Proportions are very similar, with some discrepancies seen in *Data analysis, Capturing data* (fewer in SAMs), and *Data quality* (higher percentage in SAMs).

Research question 2: How do questions addressing particular skills perform?

To investigate how questions addressing particular skills perform, we analysed item performance using the three measures described in Table 2 for each skill and compared them to each other. As multiple skills were often assigned to an item, we used the primary skill assigned for

Table 2: Measures of item performance

Measure	*Description*
Facility	The mean mark on the item divided by the maximum mark was used to measure the difficulty of an item.
Omit	The proportion of students that did not attempt the item.
R_rest	The correlation between the item mark and the paper total excluding the item. R_rest is used to evidence whether items discriminate between good and weak candidates, and positive values indicate that pupils doing well on the question also do well overall.

each item. We used the facility, omit, and R_rest values calculated for all items that were assigned a practical skill from both ATP and embedded IAPS papers. SAMs were not used, as item level data is not available for these papers.

To compare how well items targetting each skill performed, the mean and standard deviation of facility, omit, and R_rest values of all items assigned to each skill was calculated. Figure 7 shows the mean performance of each skill, in terms of facility (Figure 7a), omit rate (Figure 7b), and R_rest (Figure 7c), demonstrating variation in facility and omit values depending on the skill assessed. The facility and omit values suggest two broad groups of skills that vary by their level of difficulty. Here, the skills *Making measurements*, *Drawing (Biology)*, *Capturing data*, *Data handling*, and *Use of apparatus and techniques* have the highest mean facility values and the lowest mean omit values of the skills, indicating that items assigned as assessing these skills are easier compared to other practical skills. In contrast, the skills *Data Analysis*, *Data interpretation/Identifying trends*, *Data quality*, *Experimental design* and *Predicting outcomes* have the lowest mean facility values, and the highest mean omit values of the skills, indicating that items assigned as assessing these skills are more difficult compared to other practical skills. Furthermore, the items assigned as assessing *Recall* and *Diagrams and circuits* do not fit well with either group based on their mean facility and omit values, and so are assigned to neither group. Firstly, for the skill Diagrams and circuits, this may be due to the high variation and a very small number of items assigned as assessing these skills. Next, for the skill Recall, this may be due to the broad range of questions in which this type of skill is likely to be assessed. Finally, there is little variation in the mean R_rest values of each skill, indicating that all skills are performing

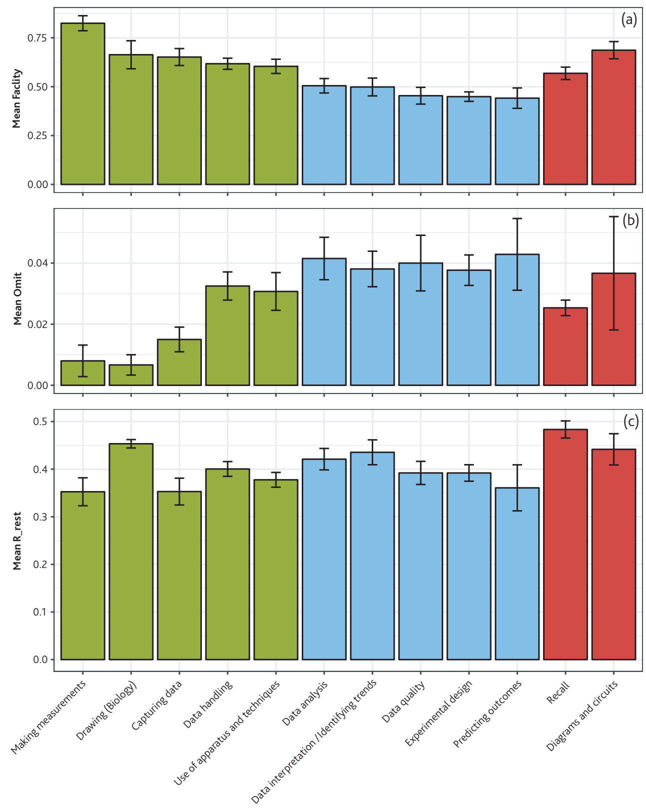


Figure 7: Measures of how well items assessing different skills functioned by subject, using both Cambridge International and OCR data
(Note: Mean facility values (a), mean omit values (b), and mean R_rest values (c), with standard error bars. The groups are distinguished by colour whereby the group with high facility and low omit are in green, the group with low facility and high omit are in blue, and the skills that do not fit in either group are in red).

to a similar level in terms of distinguishing between candidates. While the item performance measures varied between subjects (Biology, Chemistry, Physics), there was no consistent pattern whereby certain skills in certain subjects vary in comparison to other subjects or skills.

Discussion

The work described here constitutes an attempt to develop and apply a taxonomy of written questions about practical work. Locating practical science skills within the context of an explicit framework affords a more systematic and coherent approach to classifying and conceptualising such skills in written exam papers.

The integrity of a practical science test – irrespective of whether it attempts to assess practical skills *directly*, or whether such skills are *indirectly inferred* (as they might be in a written exam or through some other secondary form of assessment) – depends to a large degree upon a comprehensible understanding and articulation of the underlying construct(s) which it seeks to characterise. If these constructs are not well defined, then it will be difficult to support the claims a test developer may wish to make about the utility of the tests, including claims that the science tests do not suffer from factors such as *Construct underrepresentation and Construct irrelevance* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p.63).

A useful (and necessary) distinction should be made between what is easy to assess, and what is important to assess. It might be argued, for example, that data skills (*Capturing data, Data analysis, Data handling, and Data quality*) are easily assessed whilst also being considered crucial to the area of practical science work. Other skills subsumed under the classification *Conceptual understanding*, are often considered an important reason for doing practical work, though interestingly the literature on the efficacy of science practical work in augmenting the development of conceptual understanding is somewhat mixed (Hewson & Hewson, 1983; Hofstein & Lunetta, 1982; Lazarowitz & Tamir, 1994; Mulopo & Fowler, 1987).

Abrahams and Reiss (2015) state that indirect assessment of practical science is more appropriate for determining a student's understanding of a skill or progress, whereas direct assessment is more appropriate to determine a student's competency. The implication being that there is a potential danger that understanding of skills that are easy to assess are assessed frequently, at the expense of the understanding of other skills, and at the expense of competency. If this is the case, depending on the purpose of the assessment, an alternative form of assessment might need to be used in combination with the written exam.

How can the taxonomy be used, and by whom?

We believe that the taxonomy can be used in a variety of ways and by a range of educational practitioners (such as teachers, AOs, curriculum developers and test developers) to:

- provide a structure for classifying established, predetermined categories of indirect practical science skills that can be used by AOs and test developers for considering their intentions with respect to the assessment claims they wish to make;
- prompt an evaluation (on the part of the test developer) of how effectively the assessment claims have been met;

- allow test developers to construct question papers that elicit a range and balance of appropriate, effective practical science skills;
- ensure an appropriate level of predictability for those writing questions (both at question and paper level);
- monitor question papers over time;
- generate, from a regulatory perspective, SAMs representative of future live papers;
- afford an opportunity for test developers to consider their intentions with respect to comparability both *within* qualifications (e.g., across science subjects) and *between* qualifications (e.g., GCSE and its international counterpart);
- enable the efficacy of formative tasks to be determined (and evaluated) in relation to the purposes and objectives of the teacher; and
- offer test design and development practitioners a means for evaluating assessments from different (competitor) AOs.

Reflecting on how the taxonomy has been used

Since developing the taxonomy, it has been used in OCR to compare the coverage of assessment of skills across the three sciences on a routine basis. This has proved useful in that it has provided information for assessment specialists to help consider the consistency of skills coverage across the subjects. The taxonomy has also been used to compare skills coverage in the SAMs and live assessments. For example, the SAMs have recently been compared with the 2017 A Level. As the SAMS exemplify for schools the types of questions and coverage in the live assessment, it is very important that the live assessment reflects the coverage exemplified in the SAMs. In this context, the taxonomy has proved invaluable to OCR. More generally, AOs can use this information to support a good match between the SAMs and live materials, thereby supporting schools in their understanding of the expectation of the assessments.

The taxonomy has helped to establish a clear and well-articulated position on the underlying construct(s) of written practical science assessments. Having a clear understanding of how practical science constructs are conceptualised will serve to operationalise those constructs for assessment purposes in the future. As such, the taxonomy will support subsequent redevelopment and/or revision of science qualifications and provide a sound rationale for the proposed changes on construct and other grounds such as practicality, impact, validity and reliability.

Conclusion

The taxonomy provides a framework for considering which practical skills are assessed and how frequently they are assessed. However, it does not address the more fundamental question of which practical skills we should assess in a written paper, and what the relative frequency of assessment each skill should have. In this study, we have compared items used to assess two different levels (GCSE and GCE), and across Biology, Chemistry, and Physics. Should GCSE/IGCSE students be assessed on the same skills (and in the same proportions) as GCE students? The GCSE/IGCSE arguably serves a broader purpose (progression and development of scientific literacy) than the GCE, which is focused on preparation for HE. Should Biology, Chemistry, and Physics follow similar patterns, or do inter-disciplinary differences entail that different proportions are appropriate? It is beyond the scope of this article to answer these questions. However, by considering the taxonomy in relation to the purpose of a science practical assessment, we believe we have provided a tool with which to frame the debate.

Acknowledgements

We are indebted to a number of people who contributed to this study. We particularly thank Diana Marosi, formerly of Cambridge Assessment International Education, for her analytical input. We would also like to thank the assessment practitioners and teachers involved in the study for their input throughout the course of the work.

References

Abrahams, I., & Reiss, M. J. (2015). The assessment of practical skills. *School Science Review, 96*(357), 40–44.

Abrahams, I., & Reiss, M. J. (2012). Practical work: Its effectiveness in primary and secondary schools in England. *Journal of Research in Science Teaching, 49*(8), 1035–1055.

Abrahams, I., Reiss, M. J., & Sharpe, R. M. (2013). The assessment of practical work in school science. *Studies in Science Education, 49*(2), 209–251.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Hewson, M., & Hewson, P. (1983). Effect of instruction using student prior knowledge and conceptual change strategies on science learning. *Journal of Research in Science Teaching, 20*(8), 731–743.

Hofstein, A., & Lunetta, V. N. (1982). The role of the laboratory in science teaching: Neglected aspects of research. *Review of Educational Research, 52*(2), 201–218.

Kerr, J. F. (1963). *Practical work in school science*. Leicester: Leicester University Press.

Kirschner, P. A., & Meester, M. A. M. (1988). The laboratory in higher science education: Problems, premises and objectives. *Higher Education, 17*(1), 81–98.

Lazarowitz, R., & Tamir, P. (1994). Research on using laboratory instruction in science. In Gabel, D. L. (Ed.), *Handbook of research on science teaching and learning*. (94–128). New York: Macmillan.

Lunetta, V. N., Hofstein, A., & Clough, M. P. (2007). Learning and teaching in the school science laboratory: An analysis of research, theory and practice. In S. K. Abel & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 393–441). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mulopo, M. M., & Fowler, H. S. (1987). Effects of traditional and discovery instructional approaches on learning outcomes for learners of different intellectual development: A study of chemistry students in Zambia. *Journal of Research in Science Teaching, 24*(3), 217–227.

Quillin, K., & Thomas, S. (2015). Drawing-to-Learn: A Framework for Using Drawings to Promote Model-Based Reasoning in Biology. *CBE Life Science Education. 14*(1):es2.

Wilson, F., Wade, N., & Evans, S. (2016). Impact of changes to practical assessment at GCSE and A-level: the start of a longitudinal study by OCR. *School Science Review, 98*(362), 119–128.